A/B Testing

Author: Palak Sarawagi

Heterogenous Causal Effects

- The results of a treatment may differ based on the unique characteristics of individuals/entities or groups in a study. This idea is referred to as Heterogenous Causal Effects.
- So individual attributes like gender, age, income, health, or any other attribute may lead to a varied effect on different individuals who are under the same treatment.
- Therefore, it is essential to consider the heterogeneous causal effects when a treatment or a test is conducted.
- Heterogeneous causal effects are used to measure how a given treatment may affect different subjects differently. They also help us to identify which subgroup has a higher or lower effect due to a treatment.
- The model for a Heterogenous causal effect is given by: $y = \alpha + \beta T + \gamma Z + \sigma TZ + \varepsilon$
  In this model, Z determines different populations. Say in the above model $Z_i \in \{0,1\}$, then Z=0 is a set of populations based on specific characteristics, and Z=1 is another population set.
  Causal effect if Z=0: $[\partial y/\partial T \mid Z=0] = \beta$
  Causal effect if Z=1: $[\partial y/\partial T \mid Z=1] = \beta + \delta$
  Clearly, due to the contribution of Z, the causal effect of T is different in both sets of populations.

The effect of the Intention to Treat (ITT)

- Randomized Control Trials are usually performed to determine the causal effects in an accurate and unbiased manner. Based on this randomization, individuals in a group are put into Treatment or Control groups.
- Irrespective of whether a participant in the treatment or control group finished the treatment or not, data from all participants are used in the ITT analysis.
- So, if any participant decides to defy their treatment, their data is also included in the ITT analysis. This may lead to incorrect determination of causal effects.
- In other words, ITT assumes that all the people in the treatment and control group adhere to the assigned treatment. However, in reality, there can be defiers who do the opposite of what they have been asked to do.
- The model for ITT is given by: $y = \alpha + \beta T + \varepsilon$
  T is randomized here, so we measure β without bias as $Cov(T, \varepsilon) = 0$.

The Local Average Treatment Effect (LATE)

- As mentioned above, in a Randomized Control Trial, there are usually some defiers who do not adhere to the assigned treatment. So LATE measures the average impact of treatment on only a segment of people who would accept the treatment. In other words, LATE determines the causal effect of a treatment on the compliers and excludes any/all defiers.

- LATE is a better estimator for a causal effect. People in the treatment group who adhered to the intention to treat (took the treatment) are compared to those in the control group who adhered to the intention to treat (did not take the treatment).
- Here Cov (X, ε) ≠ 0. So calculating the causal effect requires further calculations.
- Mathematically, LATE is calculated as the difference in the average effect Y when X=1 and average effect Y when X=0, but only for compliers.
  LATE = [Y'|X=1, C] – [Y'|X=0, C]
  LATE= P[Compliers] * [Y'|X = 1] – [Y'|X = 0]

Examples:

(a) To anticipate the causal effect of telling people that this drug is available in the market(T) on the average level of health(y), we will run a regression model where we regress y on T. As T is randomized, Cov (T, ε) = 0.

The regression equation will look like $y = \alpha + \beta T + \varepsilon$ where Cov (T, ε) = 0.

In this case, we are not concerned with the fact that the people in the treatment and control group actually take the drug. This regression is entirely based on the **intention to treat (ITT).**

(b) To anticipate the causal effect of taking this drug once available in the market on the average level of health, we need to estimate the **Local Average Treatment Effect (LATE)**.

As the drug is available in the market, people can now self-select whether to take the drug or not (X-treatment). So we need to run a regression model where we regress y on X. However, here Cov (T, ε) ≠ 0. This regression is not straightforward.

So, the local average treatment effect here is the difference between the effect of those who actually took the drug and those who did not take the drug available on the market.

LATE = [Y'|X=1, C] – [Y'|X=0, C]   where C is compliers.

(c) To anticipate the average effect of sending emails to these subjects (where T_i is the randomized treatment) on the likelihood of buying the product (where Y_i is a 0/1 indicator for whether person i buys the product), we will run a regression model where we regress Y on T. AsT is randomized, Cov (T, ε) = 0.

The regression equation will look like $y = \alpha + \beta T + \varepsilon$ where Cov (T, ε) = 0.

In this case, we are evaluating if sending an email enables a person to buy a product. We do not consider the fact that the person actually opened and read the email. So, this regression is entirely based on the **intention to treat (ITT).**

(d) To anticipate the average effect of reading the email and going to the website (where X_i is an 0/1 indicator for whether the person i followed the link on the email and went to the website) on the likelihood of buying the product (where Y_i is a 0/1 indicator for whether person i buys the product) we need to evaluate the Local Treatment Average Effect **(LATE).**

Unlike part (a), here Cov (X, ε) ≠ 0.

LATE = [Y'|X=1, C] – [Y'|X=0, C]
LATE= P[Compliers] * [Y'|X = 1] – [Y'|X = 0]

(e) To anticipate the average effect of sending emails (T) to subjects over 20 years old on the likelihood of buying the product (y), we need to consider the heterogenous variable age (say Z) when running the regression for y on T.

If our population is divided into two subgroups where $Z_i \in \{0,1\}$, then Z=0 is a set of people with an age less than or equal to 20, and Z=1 is another population set with age over 20.

As T is randomized, Cov (T, ε) = 0.

However, as Z is not randomized, Cov (Z, ε) ≠ 0.

The model for this is based on the **intention to treat (ITT) with a heterogenous causal effect**. The equation for this model is given by: $y = \alpha + \beta T + \gamma Z + \sigma TZ + \varepsilon$.

(f) To anticipate the average effect of reading the email and going to the website (X) on the likelihood of buying the product(Y) for subjects more than 20 years old, we need to consider the heterogenous variable age (say Z) when evaluating the **Local Average Treatment Effect**.

Here, Cov (X, ε) ≠ 0 and also Cov (Z, ε) ≠ 0.

The model for this is based on the **Local Average Treatment Effect (LATE) with a heterogenous causal effect**.

The equation for this model is given by: $y = \alpha + \beta X + \gamma Z + \sigma XZ + \varepsilon$.

LATE = [Y'|X=1, C] – [Y'|X=0, C]
LATE= P[Compliers] * [Y'|X = 1] – [Y'|X = 0]